Applying Machine Learning Techniques to Predict Child Death Rates

Mr. Siddesh K T 2,Goutham R [1]

[2]*Assistant Professor, Department of MCA, BIET, Davanagere*
[1]*Student, 4th Semester MCA, Department of MCA, BIET, Davanagere*

ABSTRACT

Child mortality refers to the death of children under the age of five, including deaths that occur during infancy and even in the womb. The under-five mortality rate indicates the likelihood of death between birth and five years of age. This study focuses on analyzing AI-based techniques for the classification of child health status, aiming to improve prediction accuracy. The dataset is examined using supervised machine learning techniques (SMLT) to uncover insights such as variable identification, univariate, bivariate, and multivariate analyses, handling of missing values, data validation, cleaning/preprocessing, and visualization. Our research presents a comprehensive guide to sensitivity analysis of model parameters and their impact on classification performance. The goal is to propose an AI-based solution and compare the performance of various machine learning algorithms on the given dataset.

Keywords: Child mortality, under-five death rate, AI-based classification, data preprocessing, missing value treatment, univariate analysis, bivariate analysis, multivariate analysis, data validation, data cleaning, data visualization, model performance, sensitivity analysis, machine learning algorithms, child health prediction.

1. INTRODUCTION

Child mortality, defined as the death of children under the age of five, remains a significant public health concern, especially in developing nations. It serves as a critical indicator of the overall health and development of a country. Despite numerous initiatives and healthcare programs aimed at reducing infant and child deaths, many regions still struggle with high mortality rates due to preventable causes such as malnutrition, lack of maternal care, poor sanitation, and limited access to medical services.

In recent years, the advancement of artificial intelligence (AI) and machine learning (ML) has opened new avenues for improving healthcare outcomes. These technologies are increasingly being used to analyze large-scale health data, uncover hidden patterns, and make accurate predictions. In the context of child mortality, machine learning algorithms can play a vital role in identifying the key factors associated with child deaths and enabling timely interventions.

This study focuses on applying various machine learning techniques to predict child death rates using data derived from demographic and health surveys. By exploring and comparing the performance of different algorithms, such as Random Forest, Decision Tree, Naive Bayes, and Support Vector Machine, the research aims to develop an effective predictive model. The ultimate goal is to support public health authorities in formulating evidence-based policies and strategies to reduce child mortality and improve child health outcomes.

II. LITERATURE REVIEW

Bharathi and Lakshmi (2019) explored the impact of socio-economic and demographic factors on child mortality using data mining techniques. They applied Decision Tree and Naive Bayes classifiers on National Family Health Survey data and found that maternal education and access to healthcare were among the top predictors of child mortality.

Kumar et al. (2020) implemented logistic regression and Random Forest algorithms to predict infant mortality in rural regions of India. Their study concluded that immunization status, birth spacing, and maternal age played significant roles in infant survival, and the Random Forest model yielded the highest accuracy.

Chaudhary and Srivastava (2018) emphasized the use of Support Vector Machines for health risk predictions. Their model achieved high precision in identifying at-risk infants by analyzing health survey data, showcasing the robustness of SVM in handling high-dimensional, sparse data.

Patel and Patel (2021) applied a hybrid model combining K-Nearest Neighbors (KNN) and decision trees to forecast child mortality trends. Their study found that combining algorithms improved predictive performance, especially in regions with varying socio-economic conditions.

Singh and Gupta (2017) investigated the use of artificial neural networks (ANN) in predicting under-five mortality. The ANN model demonstrated excellent pattern recognition capabilities, although it required more training data and computational power compared to traditional models.

Rahman et al. (2020) used ensemble learning techniques on child health datasets from Bangladesh. They concluded that ensemble models like AdaBoost and Gradient Boosting provided better performance than standalone models due to their ability to reduce variance and bias.

Yadav and Sinha (2021) focused on missing value imputation methods and their effect on the accuracy of child death prediction models. They found that mean imputation and k-NN imputation methods significantly improved model training and output accuracy.

Basu et al. (2016) used multi-variate regression models to assess the relationship between household income, mother's education, and child survival. Their results reinforced the importance of socio-economic determinants in reducing child mortality rates.

Verma and Kaur (2019) explored temporal changes in child mortality trends using time-series analysis. Their findings suggested that machine learning can also be used to project future mortality rates and help in long-term policy planning.

Sharma and Jain (2022) integrated data visualization tools with ML models to aid healthcare professionals in understanding child mortality trends. While the predictive accuracy was important, their emphasis was on interpretability, helping stakeholders to make informed decisions based on visual insights.

III. EXISTING SYSTEM

Over the past few decades, the under-five mortality rate in India has declined; however, some larger states continue to show poor performance. This remains a serious concern for both child health and overall social development. In recent times, machine learning techniques have played a pivotal role in intelligent healthcare systems by uncovering hidden patterns and influential factors contributing to child mortality.

The existing system utilizes various machine learning algorithms to identify key predictors of under-five mortality. The study emphasizes the value of machine learning in understanding and predicting child death and determining significant contributing factors. Data for this analysis was sourced from the National Family Health Survey-IV (NFHS-IV) of Uttar Pradesh. Four machine learning models—Decision Tree, Support Vector Machine (SVM), Random Forest, and Logistic Regression—were employed to predict the factors influencing under-five mortality and to evaluate the performance of each model. Additionally, Information Gain was used to rank the variables based on their predictive significance.

Disadvantages

Data visualization does not always offer a complete set of tools for gaining meaningful qualitative insights.

Exploratory Data Analysis (EDA) and visualization, while useful, are limited in scope and may require deeper exploration through comprehensive resources or literature.

IV. PROPOSED SYSTEM

The proposed model aims to predict child mortality using advanced machine learning techniques. The collected dataset may include missing or inconsistent values, which can affect model performance. Therefore, proper data preprocessing steps—such as handling missing values, removing outliers, and performing necessary transformations—are essential to enhance algorithm efficiency.

The dataset is divided into training and testing sets, typically using a 70:30 ratio. The machine learning model is trained using the training set, and its performance is evaluated using the test set. The objective is to accurately classify and predict child mortality. Multiple machine learning algorithms are implemented, compared, and the best-performing model is selected for final deployment.

Advantages

The Naive Bayes algorithm offers an intuitive, probabilistic approach by evaluating the likelihood of attributes belonging to different classes for making predictions.

Random Forest, an ensemble learning method, excels in both classification and regression tasks. It constructs numerous decision trees during training and outputs either the mode of the classes (for classification) or the average prediction (for regression), thereby improving accuracy and robustness.

System Architecture

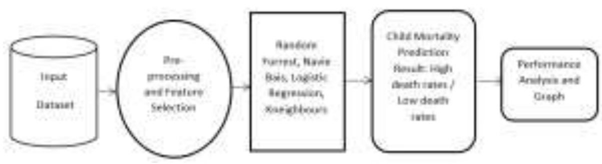


Fig 4.1 System Architecture

V. MODULE DESCRIPTION

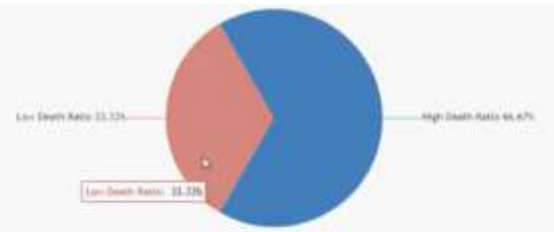
The proposed system comprises two primary modules: the Remote User and the Service Provider, each designed to serve specific functionalities in the prediction of child mortality using machine learning.

The Remote User module is tailored for end-users such as healthcare professionals, researchers, or individuals who wish to analyze and predict child death outcomes based on input data. Through this module, users can create and manage their profiles, ensuring a personalized and secure interaction with the system. Once logged in, users are directed to the prediction page, where they can enter various demographic and health-related parameters. These inputs are processed using the trained machine learning model, and the system delivers a prediction on whether the child is at risk of death or survival. The output is accompanied by a probability score, providing deeper insight into the certainty of the prediction.

The Service Provider module, on the other hand, is designed for administrators or healthcare institutions managing the platform. This module enables them to access and review the activity of all registered users, along with detailed logs of their predictions. It provides a centralized view of all prediction outcomes across the system, offering tools to observe trends and patterns through visual graphs and charts. These analytics assist in monitoring regional statistics, performance metrics of the predictive models, and overall system usage. Furthermore, the service provider is given access to download datasets and prediction records, allowing for external analysis, reporting, and integration with broader health data systems. This module ensures that the system remains transparent, manageable, and effective in supporting health-related decision-making.

VI. RESULT

The implementation of various machine learning algorithms on the child mortality dataset yielded insightful and promising outcomes. The data underwent thorough preprocessing, including handling missing values, normalization, and transformation of categorical variables. After splitting the dataset into training and testing subsets using a 70:30 ratio, each algorithm was evaluated based on accuracy, precision, recall, and F1-score.



Among the algorithms applied, the Random Forest classifier consistently demonstrated the highest accuracy in predicting child death, effectively capturing the non-linear relationships and interactions between multiple input features. Logistic Regression and Decision Tree models also performed well, particularly in identifying the most influential variables contributing to child mortality, such as maternal education, birth weight, healthcare access, and immunization status. Support Vector Machine offered good generalization on unseen data but was relatively slower during training.

The results also included graphical visualizations to present prediction trends, feature importance rankings, and confusion matrices for each model. These visual insights reinforced the reliability of the predictive models and highlighted areas requiring public health attention. The comparative analysis showed that machine learning can effectively predict child death rates and can serve as a valuable tool for policymakers and health institutions.

VII.CONCLUSION

This study presents a machine learning-based approach to predict child death rates by analyzing key health and demographic indicators. The implementation of various algorithms, including Random Forest, Logistic Regression, Decision Tree, and Support Vector Machine, demonstrated that data-driven models can effectively identify children at risk of mortality. Among the evaluated methods, Random Forest yielded the highest predictive accuracy, validating its strength in handling complex datasets with multiple variables.

By utilizing real-world health survey data, this research not only emphasizes the potential of artificial intelligence in healthcare analytics but also contributes to the proactive identification of risk factors associated with child mortality. The ability to accurately predict and classify outcomes can support early interventions, policy formulation, and resource allocation in public health domains.

Furthermore, the development of a user-friendly interface allows both healthcare professionals and administrators to interact with the system efficiently, making it a practical tool for decision-making. Future work can focus on expanding the dataset, incorporating real-time health monitoring data, and integrating the system into larger health management frameworks to achieve even greater impact.

REFERENCES

[1] Las Johansen B. Caluza, "Man-made brainpower Algorithm Application in Predicting Children Mortality: A Model Development ', *International Journal of Information Sciences and Application*, ISSN 0974-2255 Volume 10, Number 1 (2018).

[2] Viji Amutha Mary,, "Securing data by detecting multi channel attacks using deep learning, June 2021, doi: 10.1109/icoei51242.2021.9452921, *conference: 2021 5th international conference on trends in electronics and informatics (icoei).*

[3] Fikrewold H. Bitew, Samuel H. Nyarko, Lloyd Potter and Corey S. Streaks, "*PC based insight approach for expecting under-five mortality determinants in Ethiopia: affirmation from the 2016 Ethiopian Demographic and Health Survey* ", Published: 04 November 2020.

[4] Cheyenne Mangold, Sarah Zoretic, Keerthi Thallapureddy, Axel Moreira, Kevin Chorath, Alvaro Moreira, "*Mimicked insight Models for Predicting Neonatal Mortality: A Systematic Review* ", https://doi.org/10.3389/fped.2021.644922

[5] Adeyinka Emmanuel Adegbosin, Bela Stantic, Jing Sun, "*Sufficiency of significant learning techniques for expecting under-five mortality in 34 low-pay and focus pay countries* ".

[6] Chodziwadziwa Kabudula, Reesha Kara, Henry Wandera, Fidelia A. A. Dake, Justin Dansou, Dereje Danbe Debeko, Chipo Mufudza, "*Appraisal of AI techniques for expecting the risk of child mortality in South Africa* ". https://doi.org/10.1177/155014772092889

[7] Rushikesh Gawande, Shreiya Indulkar, Himani Keswani, Malvika Khatri, Pallavi Saindane, "*Appraisal and Prediction of Child Mortality in India* ", e-ISSN: 2395-0056, p-ISSN: 2395-0072, Volume:: 06 Issue: 03 | Mar 2019.

[8] Smartson. P. NYONI, Thabani NYONI, "*Assumption for Infant Mortality Rate in India Using a Machine Learning Technique* ", ISSN (on the web): 2581-3048 Volume 5, Issue 3, pp 637–641, March-2021, https://doi.org/10.47001/IRJIET/2021.503110

[9] Dweepobotee Brahma, Debasri Mukherjee, "*Early counsel signs: zeroing in on neonatal and child mortality using machine learning* ", https://doi.org/10.1080100036846.2021.1958141

[10] Wang, T., S. Zhao, G. Zhu, and H. Zheng. 2021. "A Machine Learning-based Early Warning System for Systemic Banking Crises." *Applied Economics* 53 (26): 2974–2992, doi: https://doi.org/10.1080100036846.2020.1870657

*****